IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

FILING OF A UNITED STATES PATENT APPLICATION

A GRAPHICS PROCESSING ARCHITECTURE EMPLOYING A UNIFIED SHADER

INVENTORS:

Steven Morein 10 Magazine, Apt. 801 Cambridge, Massachusetts 02139	Laurent Lefebvre 124 Parenchere Lachenaie Quebec Canada J6W 6A5
Andy Gruber 215 Pleasant Street Arlington, Massachusetts 02476	Andi Skende 49 Sheridan Drive, #11 Shrewsbury, Massachusetts 01545

ASSIGNEE:

ATI Technologies, Inc.
1 Commerce Valley Drive East
Markham, Ontario
Canada L3T 7X6

ATTORNEY OF RECORD:
CHRISTOPHER J. RECKAMP
REGISTRATION NO. 34,414
VEDDER, PRICE, KAUFMAN & KAMMHOLZ, P.C.
222 NORTH LASALLE STREET, SUITE 2600
CHICAGO, ILLINOIS 60601
PHONE (312) 609-7500
FAX (312) 609-5005

A GRAPHICS PROCESSING ARCHITECTURE EMPLOYING A UNIFIED SHADER

FIELD OF THE INVENTION

[0001] The present invention generally relates to graphics processors and, more particularly, to a graphics processor architecture employing a single shader.

BACKGROUND OF THE INVENTION

[0002] In computer graphics applications, complex shapes and structures are formed through the sampling, interconnection and rendering of more simple objects, referred to as primitives. An example of such a primitive is a triangle, or other suitable polygon. These primitives, in turn, are formed by the interconnection of individual pixels. Color and texture are then applied to the individual pixels that comprise the shape based on their location within the primitive and the primitives orientation with respect to the generated shape; thereby generating the object that is rendered to a corresponding display for subsequent viewing.

[0003] The interconnection of primitives and the application of color and textures to generated shapes are generally performed by a graphics processor. Conventional graphics processors include a series of shaders that specify how and with what corresponding attributes, a final image is drawn on a screen, or suitable display device. As illustrated in FIG. 1, a conventional shader 10 can be represented as a processing block 12 that accepts a plurality of bits of input data, such as, for example, object shape data (14) in object space (x,y,z); material properties of the object, such as color (16); texture information (18); luminance information (20); and viewing angle information (22)

1

and provides output data (28) representing the object with texture and other appearance properties applied thereto (x′, y′, z′).

[0004] In exemplary fashion, as illustrated in FIGS. 2A-2B, the shader accepts the vertex coordinate data representing cube 30 (FIG. 2A) as inputs and provides data representing, for example, a perspectively corrected view of the cube 30′ (FIG. 2B) as an output. The corrected view may be provided, for example, by applying an appropriate transformation matrix to the data representing the initial cube 30. More specifically, the representation illustrated in FIG. 2B is provided by a vertex shader that accepts as inputs the data representing, for example, vertices V_X, V_Y and V_Z, among others of cube 30 and providing angularly oriented vertices $V_{X'}$, $V_{Y'}$ and $V_{Z'}$, including any appearance attributes of corresponding cube 30′.

[0005] In addition to the vertex shader discussed above, a shader processing block that operates on the pixel level, referred to as a pixel shader is also used when generating an object for display. Generally, the pixel shader provides the color value associated with each pixel of a rendered object. Conventionally, both the vertex shader and pixel shader are separate components that are configured to perform only a single transformation or operation. Thus, in order to perform a position and a texture transformation of an input, at least two shading operations and hence, at least two shaders, need to be employed. Conventional graphics processors require the use of both a vertex shader and a pixel shader in order to generate an object. Because both types of shaders are required, known graphics processors are relatively large in size, with most of the real estate being taken up by the vertex and pixel shaders.

2

[0006] In addition to the real estate penalty associated with conventional graphics processors, there is also a corresponding performance penalty associated therewith. In conventional graphics processors, the vertex shader and the pixel shader are juxtaposed in a sequential, pipelined fashion, with the vertex shader being positioned before and operating on vertex data before the pixel shader can operate on individual pixel data.

[0007] Thus, there is a need for an improved graphics processor employing a shader that is both space efficient and computationally effective.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The present invention and the associated advantages and features thereof, will become better understood and appreciated upon review of the following detailed description of the invention, taken in conjunction with the following drawings, where like numerals represent like elements, in which:

[0009]   FIG. 1 is a schematic block diagram of a conventional shader;

[00010]   FIGS. 2A-2B are graphical representations of the operations performed by the shader illustrated in FIG. 1;

[00011]   FIG. 3 is a schematic block diagram of a conventional graphics processor architecture;

[00012]   FIG. 4A is a schematic block diagram of a graphics processor architecture according to the present invention;

3

[00013]   FIG. 4B is a schematic block diagram of an optional input component to the graphics processor according to an alternate embodiment of the present invention; and

[00014]   FIG. 5 is an exploded schematic block diagram of the unified shader employed in the graphics processor illustrated in FIG. 4A.

DETAILED DESCRIPTION OF THE INVENTION

[00015]   Briefly stated, the present invention is directed to a graphics processor that employs a unified shader that is capable of performing both the vertex operations and the pixel operations in a space saving and computationally efficient manner. In an exemplary embodiment, a graphics processor according to the present invention includes an arbiter circuit for selecting one of a plurality of inputs for processing in response to a control signal; and a shader, coupled to the arbiter, operative to process the selected one of the plurality of inputs, the shader including means for performing vertex operations and pixel operations, and wherein the shader performs one of the vertex operations or pixel operations based on the selected one of the plurality of inputs.

[00016]   The shader includes a general purpose register block for storing at least the plurality of selected inputs, a sequencer for storing logical and arithmetic instructions that are used to perform vertex and pixel manipulation operations and a processor capable of executing both floating point arithmetic and logical operations on the selected inputs according to the instructions maintained in the sequencer. The shader of the present invention is referred to as a "unified" shader because it is configured to perform both vertex and pixel operations. By employing the unified shader of the present invention,

4

the associated graphics processor is more space efficient than conventional graphics processors because the unified shader takes up less real estate than the conventional multi-shader processor architecture.

[00017]   In addition, according to the present invention, the unified shader is more computationally efficient because it allows the shader to be flexibly allocated to pixels or vertices based on workload.

[00018]   Referring now to FIG. 3, illustrated therein is a graphics processor incorporating a conventional pipeline architecture. As shown, the graphics processor 40 includes a vertex fetch block 42 which receives vertex information relating to a primitive to be rendered from an off-chip memory 55 on line 41. The fetched vertex data is then transmitted to a vertex cache 44 for storage on line 43. Upon request, the vertex data maintained in the vertex cache 44 is transmitted to a vertex shader 46 on line 45. As discussed above, an example of the information that is requested by and transmitted to the vertex shader 46 includes the object shape, material properties (e.g. color), texture information, and viewing angle. Generally, the vertex shader 46 is a programmable mechanism which applies a transformation position matrix to the input position information (obtained from the vertex cache 44), thereby providing data representing a perspectively corrected image of the object to be rendered, along with any texture or color coordinates thereof.

[00019]   After performing the transformation operation, the data representing the transformed vertices are then provided to a vertex store 48 on line 47. The vertex store 48 then transmits the modified vertex information contained therein to a primitive

5

assembly block 50 on line 49. The primitive assembly block 50 assembles, or converts, the input vertex information into a plurality of primitives to be subsequently processed. Suitable methods of assembling the input vertex information into primitives is known in the art and will not be discussed in greater detail here. The assembled primitives are then transmitted to a rasterization engine 52, which converts the previously assembled primitives into pixel data through a process referred to as walking. The resulting pixel data is then transmitted to a pixel shader 54 on line 53.

[00020]   The pixel shader 54 generates the color and additional appearance attributes that are to be applied to a given pixel, and applies the appearance attributes to the respective pixels. In addition, the pixel shader 54 is capable of fetching texture data from a texture map 57 as indexed by the pixel data from the rasterization engine 52 by transmitting such information on line 55 to the texture map. The requested texture data is then transmitted back from the texture map 57 on line 57′ and stored in a texture cache 56 before being routed to the pixel shader on line 58. Once the texture data has been received, the pixel shader 54 then performs specified logical or arithmetic operations on the received texture data to generate the pixel color or other appearance attribute of interest. The generated pixel appearance attribute is then combined with a base color, as provided by the rasterization engine on line 53, to thereby provide a pixel color to the pixel corresponding at the position of interest. The pixel appearance attribute present on line 59 is then transmitted to post raster processing blocks (not shown).

[00021]   As described above, the conventional graphics processor 40 requires the use of two separate shaders: a vertex shader 46 and a pixel shader 54. A drawback associated with such an architecture is that the overall footprint of the graphics processor

6

is relatively large as the two shaders take up a large amount of real estate. Another drawback associated with conventional graphics processor architectures is that can exhibit poor computational efficiency.

[00022] Referring now to FIG. 4A, in an exemplary embodiment, the graphics processor 60 of the present invention includes a multiplexer 66 having vertex (e.g. indices) data provided at a first input thereto and interpolated pixel parameter (e.g. position) data and attribute data from a rasterization engine 74 provided at a second input. A control signal generated by an arbiter 64 is transmitted to the multiplexer 66 on line 63. The arbiter 64 determines which of the two inputs to the multiplexer 66 is transmitted to a unified shader 62 for further processing. The arbitration scheme employed by the arbiter 64 is as follows: the vertex data on the first input of the multiplexer 66 is transmitted to the unified shader 62 on line 65 if there is enough resources available in the unified shader to operate on the vertex data; otherwise, the interpolated pixel parameter data present on the second input will be passed to the unified shader 62 for further processing.

[00023] Referring briefly to FIG. 5, the unified shader 62 will now be described. As illustrated, the unified shader 62 includes a general purpose register block 92, a plurality of source registers: including source register A 93, source register B 95, and source register C 97, a processor (e.g. CPU) 96 and a sequencer 99. The general purpose register block 92 includes sixty four registers, or available entries, for storing the information transmitted from the multiplexer 66 on line 65 or any other information to be maintained within the unified shader. The data present in the general purpose register block 92 is transmitted to the plurality of source registers via line 109.

7

[00024] The processor 96 may be comprised of a dedicated piece of hardware or can be configured as part of a general purpose computing device (i.e. personal computer). In an exemplary embodiment, the processor 96 is adapted to perform 32-bit floating point arithmetic operations as well as a complete series of logical operations on corresponding operands. As shown, the processor is logically partitioned into two sections. Section 96 is configured to execute, for example, the 32-bit floating point arithmetic operations of the unified shader. The second section, 96A, is configured to perform scaler operations (e.g. log, exponent, reciprocal square root) of the unified shader.

[00025] The sequencer 99 includes constants block 91 and an instruction store 98. The constants block 91 contains, for example, the several transformation matrices used in connection with vertex manipulation operations. The instruction store 98 contains the necessary instructions that are executed by the processor 96 in order to perform the respective arithmetic and logic operations on the data maintained in the general purpose register block 92 as provided by the source registers 93-95. The instruction store 98 further includes memory fetch instructions that, when executed, causes the unified shader 62 to fetch texture and other types of data, from memory 82 (FIG. 4A). In operation, the sequencer 99 determines whether the next instruction to be executed (from the instruction store 98) is an arithmetic or logical instruction or a memory (e.g. texture fetch) instruction. If the next instruction is a memory instruction or request, the sequencer 99 sends the request to a fetch block (not shown) which retrieves the required information from memory 82 (FIG. 4A). The retrieved information is then transmitted to the sequencer 99, through the vertex texture cache 68 (FIG. 4A) as described in greater detail below.

8

[00026]   If the next instruction to be executed is an arithmetic or logical instruction, the sequencer 99 causes the appropriate operands to be transferred from the general purpose register block 92 into the appropriate source registers (93, 95, 97) for execution, and an appropriate signal is sent to the processor 96 on line 101 indicating what operation or series of operations are to be executed on the several operands present in the source registers. At this point, the processor 96 executes the instructions on the operands present in the source registers and provides the result on line 85. The information present on line 85 may be transmitted back to the general purpose register block 92 for storage, or transmitted to succeeding components of the graphics processor 60.

[00027]   As discussed above, the instruction store 98 maintains both vertex manipulation instructions and pixel manipulation instructions. Therefore, the unified shader 99 of the present invention is able to perform both vertex and pixel operations, as well as execute memory fetch operations. As such, the unified shader 62 of the present invention is able to perform both the vertex shading and pixel shading operations on data in the context of a graphics controller based on information passed from the multiplexer. By being adapted to perform memory fetches, the unified shader of the present invention is able to perform additional processes that conventional vertex shaders cannot perform; while at the same time, perform pixel operations.

[00028]   The unified shader 62 has ability to simultaneously perform vertex manipulation operations and pixel manipulation operations at various degrees of completion by being able to freely switch between such programs or instructions, maintained in the instruction store 98, very quickly. In application, vertex data to be processed is transmitted into the general purpose register block 92 from multiplexer 66.

9

The instruction store 98 then passes the corresponding control signals to the processor 96 on line 101 to perform such vertex operations. However, if the general purpose register block 92 does not have enough available space therein to store the incoming vertex data, such information will not be transmitted as the arbitration scheme of the arbiter 64 is not satisfied. In this manner, any pixel calculation operations that are to be, or are currently being, performed by the processor 96 are continued, based on the instructions maintained in the instruction store 98, until enough registers within the general purpose register block 92 become available. Thus, through the sharing of resources within the unified shader 62, processing of image data is enhanced as there is no down time associated with the processor 96.

[00029]   Referring back to FIG. 4A, the graphics processor 60 further includes a cache block 70, including a parameter cache 70A and a position cache 70B which accepts the pixel based output of the unified shader 62 on line 85 and stores the respective pixel parameter and position information in the corresponding cache. The pixel information present in the cache block 70 is then transmitted to the primitive assembly block 72 on line 71. The primitive assembly block 72 is responsible for assembling the information transmitted thereto from the cache block 70 into a series of triangles, or other suitable primitives, for further processing. The assembled primitives are then transmitted on line 73 to rasterization engine block 74, where the transmitted primitives are then converted into individual pixel data information through a walking process, or any other suitable pixel generation process. The resulting pixel data from the rasterization engine block 74 is the interpolated pixel parameter data that is transmitted to the second input of the multiplexer 66 on line 75.

10

[00030]   In those situations when vertex data is transmitted to the unified shader 62

through the multiplexer 66, the resulting vertex data generated by the processor 96, is

transmitted to a render back end block 76 which converts the resulting vertex data into at

least one of several formats suitable for later display on display device 84. For example,

if a stained glass appearance effect is to be applied to an image, the information

corresponding to such appearance effect is associated with the appropriate position data

by the render back end 76. The information from the render back end 76 is then

transmitted to memory 82 and a display controller line 80 via memory controller 78.

Such appropriately formatted information is then transmitted on line 83 for presentation

on display device 84.

[00031]   Referring now to FIG. 4B, shown therein is a vertex block 61 which is

used to provide the vertex information at the first input of the multiplexer 66 according to

an alternate embodiment of the present invention. The vertex block 61 includes a vertex

fetch block 61A which is responsible for retrieving vertex information from memory 82,

if requested, and transmitting that vertex information into the vertex cache 61B. The

information stored in the vertex cache 61B comprises the vertex information that is

coupled to the first input of multiplexer 66.

[00032]   As discussed above, the graphics processor 60 of the present invention

incorporates a unified shader 62 which is capable of performing both vertex manipulation

operations and pixel manipulation operations based on the instructions stored in the

instruction store 98. In this fashion, the graphics processor 60 of the present invention

takes up less real estate than conventional graphics processors as separate vertex shaders

and pixel shaders are no longer required. In addition, as the unified shader 62 is capable

of alternating between performing vertex manipulation operations and pixel manipulation operations, graphics processing efficiency is enhanced as one type of data operations is not dependent upon another type of data operations. Therefore, any performance penalties experienced as a result of dependent operations in conventional graphics processors are overcome.

[00033]   The above detailed description of the present invention and the examples described therein have been presented for the purposes of illustration and description. It is therefore contemplated that the present invention cover any and all modifications, variations and equivalents that fall within the spirit and scope of the basic underlying principles disclosed and claimed herein.